Technical Research Note 200

AD

CERTITUDE JUDGMENTS IN AN OPERATIONAL ENVIRONMENT

James D. Baker
U. S. Army Behavioral Science Research Laboratory

and

James M. McKendry and Douglas J. Mace
HRB-Singer, Inc.

BESRL

U. S. Army
Behavioral Science Research Laboratory

November 1968

32

Technical Research Note 200

AD

CERTITUDE JUDGMENTS IN AN OPERATIONAL ENVIRONMENT

James D. Baker
U. S. Army Behavioral Science Research Laboratory

and

James M. McKendry and Douglas J. Mace
HRB-Singer, Inc.

Seymour Ringel, Task Leader

SUPPORT SYSTEMS RESEARCH DIVISION
Joseph Zeidner, Chief

U. S. ARMY BEHAVIORAL SCIENCE RESEARCH LABORATORY

Office, Chief of Research and Development
Department of the Army

Room 239, The Commonwealth Building
1320 Wilson Boulevard, Arlington, Virginia 22209

November 1968

Army Project Number
2J062105A723

Contract No. DA-49-092 ARO-19
Tactical Operations System b-11

BESRL Technical Research Reports and Technical Research Notes are intended for sponsors of R&D tasks and other research and military agencies. Any findings ready for implementation at the time of publication are presented in the latter part of the Brief. Upon completion of a major phase of the task, formal recommendations for official action normally are conveyed to appropriate military agencies by briefing or Disposition Form.

FOREWORD

Technological advancements have led to increased speed, mobility, and destructive power of military operations. To permit commanders to make tactical decisions consistent with rapid change and succession of events, information on military operations must be processed and used more effectively than ever before. To meet this need, the Army is developing automated systems for receipt, processing, storage, retrieval, and display of different types and vast amounts of military data. There is a concomitant requirement for research to determine how human abilities can be utilized to enable the command information processing systems to function with maximum effectiveness.

One objective of the COMMAND SYSTEMS Program is to provide research information by which decision making and information assimilation from displays may be facilitated. To this end, studies are conducted on such information presentation factors as amount, density, format, coding, specificity-generality, alpha-numeric vs symbolic displays, probability data and certitude, rate and degree of updating, individual and group displays, and the relative utility of different sensory and display modalities. The entire research effort is responsive to requirements of RDT&E Project 2J062106A723. "Human Performance in Military Systems," and to special requirements of the Assistant Chief of Staff for Force Development, the Assistant Chief of Staff for Intelligence, and the Automatic Data Field Systems Command.

In September 1967, the Behavioral Science Research Laboratory established a Command Systems Field Branch in Europe to conduct human performance research in connection with the evaluation of an experimental Tactical Operations System. The present publication describes one of the first studies conducted by the Field Branch. The study deals with certitude judgments accompanying intelligence data obtained in a large-scale field exercise.

J. E. UHLANER, Director
U. S. Army Behavioral Science
Research Laboratory

CERTITUDE JUDGMENTS IN AN OPERATIONAL ENVIRONMENT

BRIEF

Requirement:

To analyze certitude judgments accompanying intelligence spot reports received during an Army field exercise in Europe; particularly, to determine how quantity and quality of such reports would have been affected if a computerized tactical operations system (TOS) had been in use during the field exercise.

Procedure:

In the intelligence cycle, G2 spot reports include subjective evaluations of the information expressed in a standard rating format. All messages (N = 2039) filed by two divisions of one corps during the seven days of the field exercise were examined for presence of the required ratings of reliability of the information source and accuracy of the information. Assessment was made of the effect the TOS would have had in increasing quantity of evaluations. Quality of the evaluations was also assessed.

Findings:

1. Intelligence (G2) spot reports constituted 70 percent of all messages.

2. Half the spot reports lacked the required source reliability and accuracy ratings. Some omissions could be attributed to unavailability of standard spot report forms (which provide spaces for the ratings).

3. Reliability and accuracy ratings were not independent. The distributions revealed that senders tended to assign high scale values to ratings.

4. Investigation of the rating scales themselves indicated that the source reliability scale was unidimensional, whereas the accuracy scale was multidimensional.

Utilizations of Findings:

Introduction of the TOS may substantially increase the quantity of the certitude data available for decision making. The present rating scales should be re-evaluated in an attempt to improve rating quality.

CONTENTS

TABLES

FIGURES

CERTITUDE JUDGMENTS IN AN OPERATIONAL ENVIRONMENT

The COMMAND SYSTEMS Program within the U. S. Army Behavioral Science Research Laboratory (BESRL) is concerned with problems of information presentation, processing, and utilization in command and control systems. Specific aspects of information processing systems are examined with respect to requirements for handling information and its presentation to the human users of the system.

Coincidental with BESRL research interests, the Automatic Data Field Systems Command, Headquarters U. S. Army Europe, and Seventh Army have initiated plans for the design verification of an automated tactical operations system (TOS). The TOS automates certain intelligence and operations functions which the commander may monitor and use to direct tactical operations.

To provide for the melding of support for both BESRL research and TOS development, BESRL has established a Command Systems Field Branch within USAREUR. Research activities of the branch are directed to support of on-the-spot development of human factors aspects of the TOS. As an added dividend, BESRL's in-house research programs are enriched through infusion of operationally oriented concepts and close association with field problems for which research is indicated.

In the present study, a specific aspect of the process of obtaining tactical information during a field exercise--namely, the evaluation of the information supplied--was examined for implications for TOS design and for BESRL research.

CONTEXT OF THE STUDY

A fundamental characteristic of any command and control system is that the men in the system deal with fallible data and use such data in making decisions and carrying out their missions. Since the data are fallible, the man in the system is constantly operating in an uncertain environment.[1]

[1] Certainty is here defined as a psychological state where a person has no doubts about the outcome of a given state of affairs. To the extent that he has doubts about the outcome, he is uncertain.

Considerable research has been conducted on the evaluation of fallible data prior to choice of a course of action. Most of the experiments have been small-scale, rigidly controlled laboratory studies. The concern has been with how well individuals can express their certitude. In many studies, each subject is instructed to assign a number between zero and one that represents the extent to which he believes certain events will occur, in other words, to state his subjective probability concerning an outcome. Edwards[2], for example, has examined this procedure for potential utilization in Air Force command and control systems. Another common procedure permits the decision maker to express his certitude, or confidence in the data, by marking some point on a graphic scale. Andrews and Ringel[3] used this technique in examining display variables in a simulated Army command and control system. However, expressions of certitude have not been studied in the practical context of an everyday working environment--perhaps because in the real world such expressions are most often handled intrinsically rather than extrinsically, the decision maker manipulating his evaluations "in his head" rather than writing them down.

In the evaluation process of the G2 operations cycle, however, expressions of certitude are made explicit. A system has been standardized for use by NATO Army Forces in which a sender reports his certitude concerning the reliability of the source and the accuracy of the information being sent. The system uses a six-point letter scale to indicate source reliability and a six-point numerical scale to indicate accuracy of the information. Figure 1 is a sample G2 spot report form. The rating scales constitute Item 5. The Appendix contains a detailed description of the intelligence cycle and the standards used in assigning the values. Labels attaching to the alpha-numeric symbols are as follows;

Reliability of source:

A. Completely reliable
B. Usually reliable
C. Fairly reliable
D. Not usually reliable
E. Unreliable
F. Reliability cannot be judged

---

[2] Edwards, W. Probabilistic information processing system for diagnosis and action selection. In Information System Sciences: Proceedings of the Second Congress. J. Spiegel and D. Walker (Eds.), Washington, D.C. Spartan Books, Inc. 1965.

[3] Andrews, R. S. and S. Ringel. Certitude judgments and accuracy of information assimilation in visual displays. BESRL Technical Research Note 145. Washington, D. C. U. S. Army Behavioral Science Research Laboratory. April 1964.

Accuracy of an item of information:

1. Confirmed by other sources
2. Probably true
3. Possibly true
4. Doubtfully true
5. Improbable
6. Truth cannot be judged

Example: Information from a fairly reliable source and considered probably true would be rated C-2.

OBJECTIVES OF THE STUDY

The major focus of the present study was the certitude judgment process within the TOS. Attempt was made to derive logically the problems which could result when the evaluation process occurs within the TOS, as well as to estimate the quantitative and qualitative improvement that the TOS could be expected to bring. The selective retrieval capability of the TOS will permit corroborating intelligence reports in ways that in the manual system entail prohibitive effort. Hence, while the TOS may produce some qualitative improvement in the intelligence cycle, major emphasis is on reducing the large amount of effort that goes into general information processing in the manual system.

In September 1967, the Seventh Army held a large-scale field exercise. During the exercise, it appeared that the reliability and accuracy evaluations were frequently being omitted from the G2 spot reports and that the ratings made were high.

What are some of the implications of these observations? From the standpoint of an automated TOS, insertion of the alpha-numeric value is mandatory; otherwise, the message is rejected by the system. To elaborate, an input message is initially processed through an editing and validation subroutine which examines each message to insure that all required items have been completed. If no entry is found in the required data field, the message is returned to the originator with an indication of the error. Therefore, it was important to ascertain what percentage of inputs would have been rejected if the TOS has been employed in the field exercise.

From the standpoint of utilization in the G2 decision process, it was important to determine if the distributions of judgments of reliability of source and accuracy of information were skewed. While an automated TOS may indirectly improve the evaluation process by increasing the amount of information available to the G2, the system does not provide direct control over the quality of the information.

_____
(CLASSIFICATION)

G2 SPOT REPORT FORM

JOURNAL ENTRY _____

FROM: _____ RECEIVED BY: _____
(Headquarters - Name) (Name)

MESSAGE NUMBER _____ DATE/TIME RECEIVED: _____

EVENT

1. WHEN: _____ 2. WHERE: _____
(Date/Time of event) (UTM Coordinates)

3. WHO/WHAT/WHY (Description of event): _____

_____

_____

4. SOURCE: _____
(Original supplier of information)

5. SENDER'S EVALUATION: A B C D E F 6. IMMINENCE: a. Positive
(Circle answer) 1 2 3 4 5 6 (Circle answer) b. Negative

7. REMARKS: _____

_____

STANDARD DISTRIBUTION:
White: Action Officer-Map-G2 Journal
Pink: OB Map-Briefing Map
Blue: G2 Air-FSCC

OTHER DISSEMINATION:

Other Headquarters:

( ) ASA	( ) TAC CP ___	( ) 7th Army _____	( ) ___ Div _____
( ) CI	( ) G1 _____	( ) CENTAG _____	( ) ___ Div _____
( ) G2 Control ___	( ) G2 _____	( ) ___ Corps _____	( ) 14th A/C ___
( ) Alt CP _____	( ) G3 _____	( ) 3d GE Corps ___	( ) 37th Engr ___
( ) Rear CP _____	( ) G4 _____	( ) ___ Corps _____	( ) 319th ASA ___

_____
(CLASSIFICATION)

Figure 1. Example of a G2 Spot Report Form

METHOD

Data Sample

For baseline comparison purposes, data on manual system operations have been obtained during field exercises. Data for the present study were collected during one such exercise. The data population consisted of headquarters records, journal logs, and spot reports from the headquarters of five divisions, two corps, and a field army. While analysis of all the data relevant to source reliability and information accuracy would have been desirable, the effort would have been a major drain on the manpower and resources available within the BESRL field unit. The decision was therefore made to take a representative sample of the data.

Determination of what constitutes a representative sample was made on the basis of the probable configuration of the first increment of the TOS. The initial hardware equipment consists of one transportable Central Computing Center (CCC), four Remote Station Data Terminals (RSDT), and 18 User Input/Output Devices (UIOD). Since the TOS will most likely be implemented in the field army by tying in one corps and two of its divisions, the sample chosen for the present analysis consisted of the G2 spot reports from two divisions, the corps to which the divisions reported, and the field army.

Certain of the sample data were excluded from the analysis, for example, classified spot reports and journal items which the examiner could not definitely identify as requiring the NATO rating. The data loss, however, was minimal (2%).

Analysis

Data were analyzed to determine the following:

1. Number of G2 spot reports generated.

2. Number of G2 spot reports containing ratings of reliability of source and accuracy of information.

3. Number of G2 spot reports containing a partial rating.

4. For G2 spot reports containing a rating, the distribution of ratings by alpha-numeric category.

RESULTS

Number of G2 Spot Reports Generated

A comparison between the total message load of the two divisions and the number of spot reports constituting the message traffic is presented in Table 1.

Table 1

DATA FROM THE TWO DIVISIONS EXAMINED

	Total Journal Entries		No. of Spot Reports		No. of Reports Rated	
	N	%	N	%	N	%
Division A	887	43.5	598	41.6	499	69.7
Division B	1152	56.5	840	58.4	217	30.3
Total	2039	100.0	1438	100.0	716	100.0

Approximately 70 percent of the messages examined were spot reports. This figure is an approximation because it does not reflect the classified and doubtful entries which were dropped from the data sample although retained in the figures for total journal entries. However, an indication of how representative of the true population the value is was obtained by comparison with a prior field exercise in which data were collected on all message traffic for one of the divisions examined in the present study. In the prior field exercise, G2 spot reports made up 62 percent of the G2 traffic. In the present exercise, G2 spot reports made up 67 percent of traffic of the same division. The difference was not statistically significant ($x^2(1) = 1.57$, $p < .30$).

Number of Spot Reports With the Required Ratings

Only half (49.7 percent) of the G2 spot reports in the sample contained evaluations concerning the reliability of source and accuracy of information. Thus, only 50 percent of the reports would have been accepted as inputs into the TOS.

Several comments are in order concerning these figures. In most instances, G2 spot reports were submitted on forms similar to that shown in Figure 1. In some cases, however, the supply of standard forms was exhausted--or for some reason was not used--and the spot reports were submitted on plain paper. None of the reports on plain paper contained the required certitude judgments. Also, many spot reports submitted by controllers in the exercises failed to include the evaluations.

In the TOS, before a message is inserted into the system it is prepared in worksheet form by the action officer. Figure 2 is an example of a message identifying an enemy unit as it would appear on the user's display. After the necessary items of information are completed by the action officer, the completed message would look like that shown in Figure 3. The NATO rating entry in this case is B-2, indicating that the information is deemed to be from a "usually reliable source" and is

"probably true". The expectation would be that the requirement to format the data for completion of the standard form would be conducive to inclusion of the information. However, even spot reports made on standard forms quite often failed to contain the certitude rating. Hence, a formatting requirement alone is not sufficient to insure completion of the rating item. In the TOS, the completed worksheet will be handed to the UIOD operator, and he will type the information for input into the system. As he does so, the same format as is used by the action officer is displayed on the UIOD. The operator's typed entries are also displayed, and the completed message would be identical to that shown in Figure 3. The UIOD operator, therefore, may catch any omission before entering the message into the system.

```
     (ENEMY UNIT IDENTIFICATION DATA)
ORIGIN/     ;SCTY/       ;RESTR/       ;
EVAL/     /   ;SOURCE/       /       /         ;
UNIT/    /      /       /       /       ;
CON-RES/ ;
ACTV/        ;
TIME/              ;
LOC/             ;
PARENT/  /         /      /       /       ;
RMKS/
                                                    ;
```

Figure 2. Example of an enemy unit identification message format as it would appear on the user's display

```
     (ENEMY UNIT IDENTIFICATION DATA)
ORIGIN/STG2;SCTY/UNCLAS;RESTR/Q      ;
EVAL/INFO/B2;SOURCE/LRRP  /OBSN  /         ;
UNIT/UR/64TH   /GDS   /INF     /DIV     ;
CON-RES/R;
ACTV/POSIT  ;
TIME/270515ZDEC63;
LOC/VR262068500;
PARENT/  /       /       /       /       ;
RMKS/CZECH UNITS APPEAR TO BE ATTACHED TO THIS DIV
                                                    ;
```

Figure 3. Completed message for enemy unit identification as it would appear on the user's display

Considering the two-stage filtering in the TOS, it seems unlikely that an omission would occur. However, even assuming an oversight on the part of the action officer and the UIOD operator, the TOS system would immediately detect the omission. Upon insertion, the editing and validation subroutine would process the message and return it to the originator with an error notification. Return is accomplished by a "split-screen" technique: The text remains displayed, but an error notification appears on the lower portion of the display. The required data would then have to be inserted by the originator before the message would be accepted for the system's data base. The immediate error-notification feedback loop just described should result in very rapid extinction of omission behavior on the part of senders. Consequently, the TOS, rather than providing certitude data from only half the spot reports received, would more likely provide certitude data from all the spot reports. Any additional evaluations received as a result of this processing would represent a gain over the present rate.

There was marked difference in the performance of the two divisions in the quantity of ratings. Division A's spot reports contained the rating 83.4 percent of the time whereas Division B's spot reports had the ratings only 25.8 percent of the time. Explanation was sought for the difference in the divisions.

If the ratings are critical to the G2 decision process, incidence of rating would be expected to increase over time as a function of frequent requests on the part of the decision maker. The data, however, do not entirely support this assumption. Figure 4 shows the percent of rated spot reports as a function of days into the exercise. It can be seen that Division A rapidly asymptotes to a high level of inclusion, and that by the end of the exercise it was exhibiting near perfect performance. Division B, on the other hand, started at approximately the same level as A and then rapidly declined to a level of 10 to 20 percent. The data were examined to determine if the low percentage of ratings in Division B could be attributed to differences in absolute load demands. Did Division B process an excessively large number of messages on the days when rating inclusion went down? When the data shown on a day-by-day basis in Figure 5 were pooled, it was found that Division B had a higher message load than Division A ($X^2 = 40.72$, $P < .001$). This difference appears to bear little relationship to the pattern of rating behavior shown in Figure 4, however. For example, during the last four days of the exercise, Division B rated only a small proportion of its reports (see Figure 4); yet, on these same days, the absolute message load was comparatively low (see Figure 5). Further, although the absolute message loads on days 1 and 6 were roughly equivalent, day 1 had approximately thirty percent more completed ratings than day 6.

The data were further examined in an attempt to discover whether the lower percentage of ratings in Division B could be attributed to differences in relative load demands: Did spot reports constitute a greater proportion of daily messages for Division B than for Division A? Figure 6 reveals that, in a relative sense, no appreciable difference existed between the two divisions in this respect.

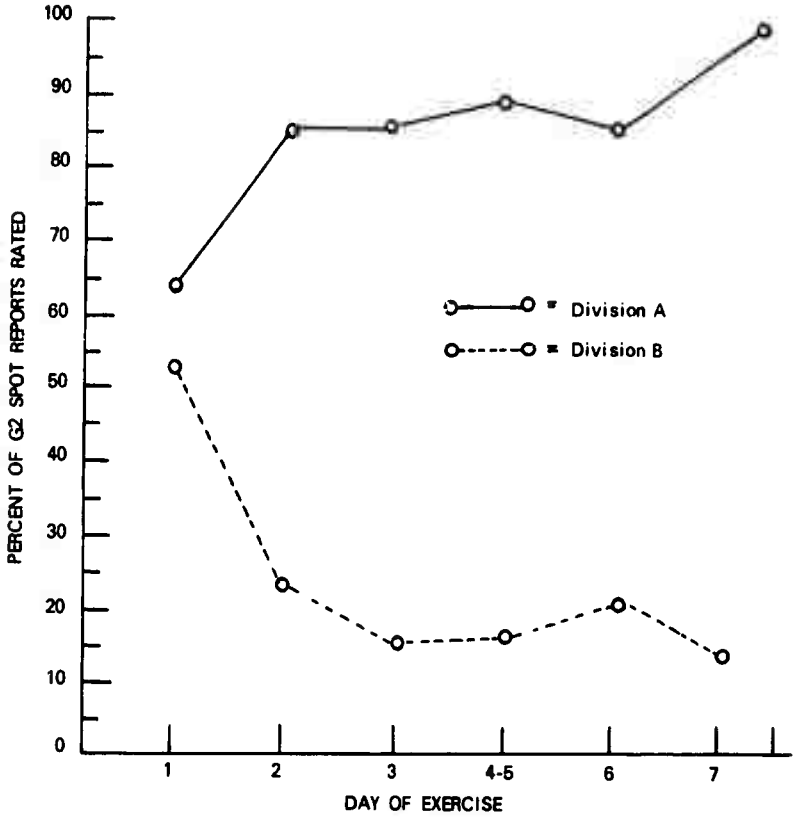


Figure 4. Percent of rated spot reports as a function of days into the exercise. Days 4 and 5 have been combined since this was the period of the low traffic (Saturday-Sunday weekend break).

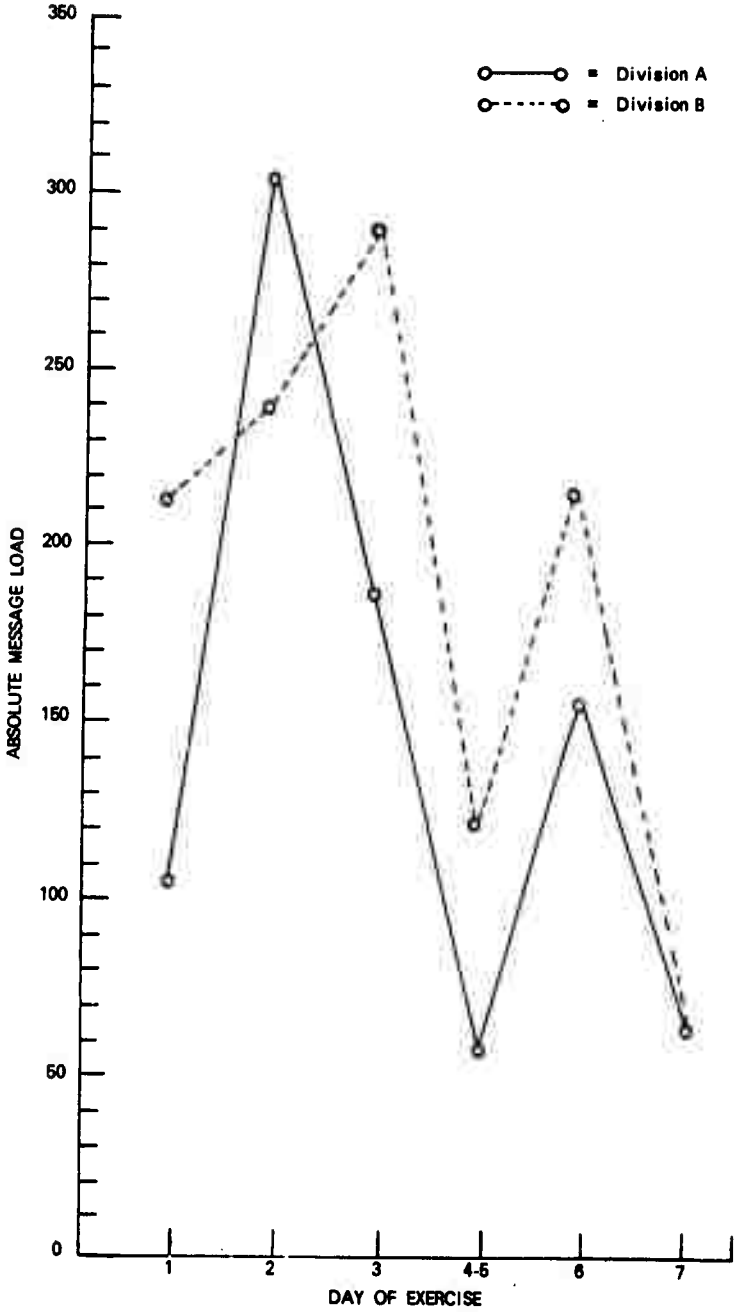


Figure 5. Absolute message load by days into the exercise

Another possibility was that one division submitted more spot reports on plain paper (ratings were invariably omitted when standard forms were not used). Approximately 99 percent of the spot reports for Division A were recorded on standard forms as opposed to approximately 66 percent for Division B. Almost half (45.8 percent) of the failures to rate in Division B, then, can be attributed to the use of nonstandard forms for reporting.[1] The implication of the TOS is obvious: Provide sufficient standard forms for the action officer and require that he use them. Granted the TOS has a built-in crutch in that the message format is displayed on the UIOD. However, the action officer would himself have to input messages on line.

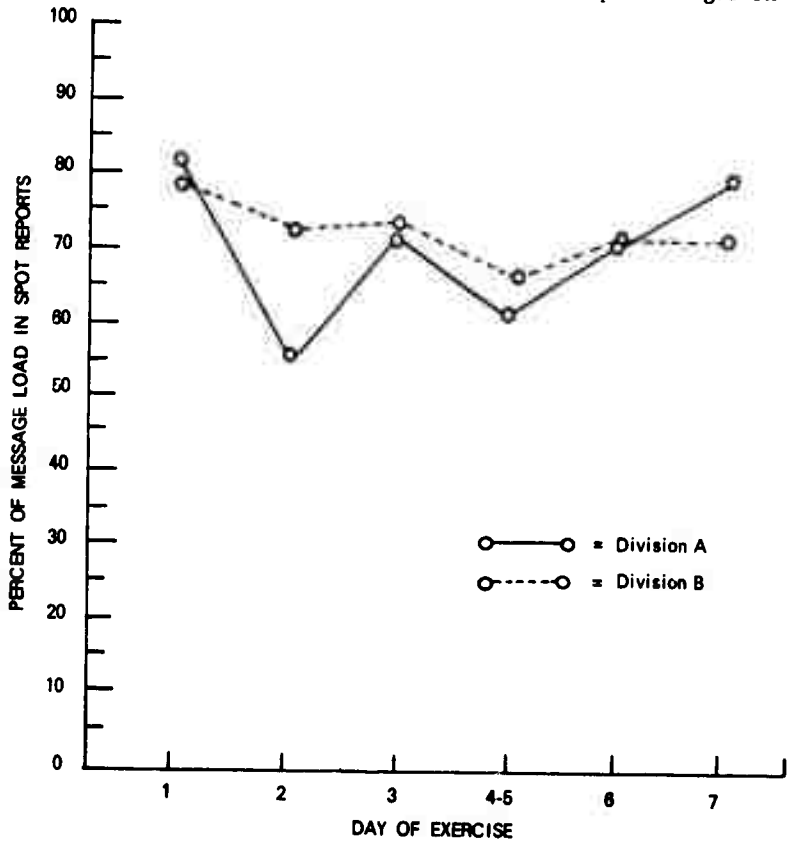


Figure 6. Percent of daily message load constituted of spot reports. Data for days 4 and 5 have been aggregated.

[1] If all reports on the standard forms had been rated, only 34 percent of the 840 reports in Division B would be nonrated, a total of 285.6. Instead, 623 reports were nonrated. The 45.8 percent figure was arrived at by the computation: 285.6/623 x 100.

A portion of the difference between the two divisions, however, was not accounted for by the difference in availability of standard forms. In the case of Division B, why did the incidence of rating not increase over time as a function of frequent requests on the part of the decision makers who would use the evaluations? Unfortunately, the raw data examined do not provide an answer to this question. In the case of Division A, a demand for the rating information may have existed, but it is equally likely that the higher proportion of rated reports was the result of rigid adherence to SOP, strict supervision, or both. Why Division A rated a higher percentage of its spot reports remains a difficult experimental question that only a controlled study can answer.

Number of G2 Spot Reports With Partial Ratings

One concern of the present study was with partial ratings, provided for by FM 30-5: "Combat Intelligence," July 1963. This concern is tied to one of the constraints that computer systems, in general, impose on human information handling tasks. In man-man systems, "maybe" events are tolerable. In a man-computer system, there is no room for "maybeness". Either the information is present or it is not. The machine makes no assumptions that "maybe the man forgot" or "maybe there is not sufficient information for making such a decision". Rather, the simplest device for machine handling of information is to make some entry mandatory; an entry such as "F" or "6" (reliability or truth cannot be judged) provides the computer with the material for completeness check.

In an automated TOS, a partial rating has the same effect as no rating: Absence of either source reliability or accuracy rating will result in rejection of the message. What, then, was the incidence of partial ratings in the field exercise under study? Of the 716 spot reports rated, 20 contained partial ratings. Of the partial ratings, 18 were on forms from which the accuracy scale values (1-6) had been omitted in reproduction. In view of the finding that total omission of the rating occurred when the spot report was submitted on plain paper, the format itself appears to have had an impact on the originator's behavior. The two remaining partial evaluations were likely a reflection of human error. Thus, partial ratings should not be a major source of difficulty in an automated TOS.

Distribution of Alpha-Numeric Ratings

Theoretically, the data concerned with the reliability of the source (alpha rating) and the accuracy of the item (numeric rating) are independent. According to FM 30-5: "Although both letters and numerals are used to indicate the evaluation of an item of information, they are independent of each other. A completely reliable agency may report information obtained from a completely reliable source which, on the basis of other information, is judged to be improbable. In such a case, the evaluation of the information is A-5. A source known to be unreliable may provide information that is confirmed by other sources and is of undoubted accuracy. In such a case, a report is evaluated E-1."

The output of the overall rating process (reliability and accuracy), then, should be a set of independent alpha-numeric values covering a wide range of expressed certitude. However, in the field exercise studied, both source reliability and accuracy ratings were most often high rather than being distributed along the scale. Shown in Table 2 is the distribution of alpha and numeric ratings in the sample. Two principal results are indicated: First, the distribution of values for each scale, indicated by row and colums totals, is far from normal. Of the 695 messages rated, 85 percent were rated B for reliability, and 76 percent received an accuracy rating of 2. Second, 87 percent of the ratings fell along the diagonal--A-1, B-2, C-3, etc. This correspondence implies that the two scales are not independent and that the information presented by one is redundant with respect to information presented by the other.

These results indicate that the rating scales were not utilized to full potential. It is recognized that this was a field exercise, and that ambiguous data or data without historical continuity may not have been introduced into the problem. It may well be that the expressed certitude values are not representative of what would be found in real-world wartime operations. It would appear that future exercises would be benefited if ambiguous data were introduced in order to afford personnel greater opportunity to exercise decision-making skill.

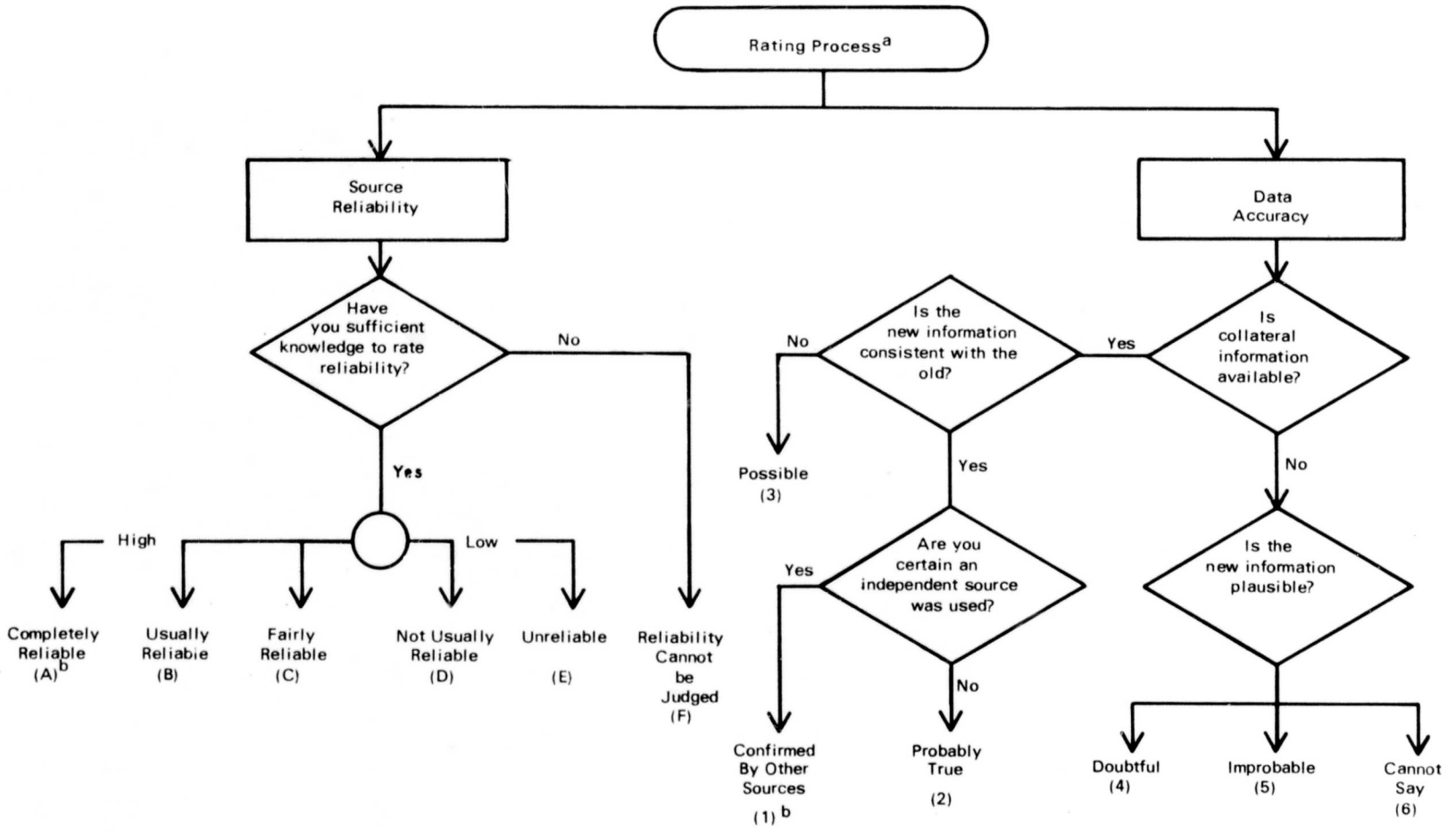
An analysis of the NATO rating scales revealed one scale (reliability) to be unidimensional and the other (accuracy) multidimensional. Figure 7 shows a decision table for use of the two scales. In rating reliability of the source, the rater selects a point along a continuum from high to low. In rating the accuracy of an item of information, the rater uses a more complex procedure. A new dimension is used at each stage in such a way that either two or three dimensions are involved in the final rating. It may be that the close correspondence between ratings (see Table 2) reflect the fact that persons often do not use tools the way the designer intended them to be used. The multidimensionality of the rating scale for measuring accuracy could certainly be responsible for the misuse of the scale. The unusual complexity of the scale could be causing messages to be incorrectly rated. If this be true, the condition could be rectified by expanding the scale into several unidimensional ratings or by providing decision aids to help the rater.

Table 2

DISTRIBUTIONS OF RATINGS OBTAINED DURING FIELD EXERCISE

Reliability of the source	Accuracy of the Information 1 Confirmed by other sources	2 Probably true	3 Possibly true	4 Doubtfully true	5 Improbable	6 Truth cannot be judged	Row Total*
A. Completely reliable	45	11	2	0	0	0	56
B. Usually reliable	11	518	57	2	0	0	588
C. Fairly reliable	0	0	5	1	1	0	7
D. Not usually reliable	0	0	0	8	0	0	8
E. Unreliable	0	0	0	0	3	1	4
F. Reliability cannot be judged	0	0	1	0	0	31	32
Column Total*	54	529	65	11	4	32	695

*The sum of the rows and of the columns (Σ_{ij}) equals 695, rather than 716, since the 20 partial ratings and 1 dual rating could not be included.



[a]Based on definitions and instructions in FM 30-5

[b]The letters in parentheses represent the value for the source reliability scale and the numbers in parentheses represent the numerical value for the data accuracy scale.

Figure 7. Decision table for determining source reliability and accuracy of information.

SUMMARY OF FINDINGS

Of the 1,438 spot reports examined, only 716, or about 50 percent, contained certitude ratings. Only 695 of these contained acceptable ratings of both source reliability and message accuracy. Although some of the omissions could be attributed to unavailability of standard spot report forms, the fact remains that in approximately half the cases certitude ratings were absent.

The incidence of partial ratings was extremely low (20 of the 716 rated messages), and the few partial ratings that did occur could be attributed to a defect in the rating forms or human error.

There was a pronounced tendency to assign high reliability ratings to sources of spot reports. More than 90 percent of the rated messages were considered to come from "completely" or "usually reliable" sources. Less than 2 percent of the messages were judged to come from a source which was "not usually reliable" or "unreliable". Ratings of information accuracy were only slightly lower. Approximately 84 percent of messages were rated as being either confirmed by other sources or probably true, the latter rating being assigned to the great bulk of all ratings (76 percent). Approximately 9 percent of the messages received a "possibly true" accuracy rating, and less than 2 percent received "doubtfully true" or "improbable" ratings. The remainder of the cases, roughly 5 percent, received a "truth cannot be judged" rating.

The interdependence of the two ratings was striking. Of the 36 reliability-accuracy combinations which could appear, less than half (15) actually appeared. A total of 7 combinations described more than 95 percent of the messages, while more than 87 percent showed exact correspondence--A-1, B-2, ...F-6.

CONCLUSIONS AND IMPLICATIONS

The results pointed up a number of conditions surrounding the certitude judgments. First, practices within a division may affect the proportion of spot reports carrying certitude ratings. Second, there is a pronounced tendency to assign high certitude ratings to spot reports. Third, the two separate certitude ratings are highly correlated. Fourth, a large number of spot reports do not include certitude ratings provided by the originator of the message.

Implications of the findings are straightforward. The formatting requirements in the TOS can be expected to standardize the rating procedures within the divisions. Since in the manual system the lack of formatted worksheets for submitting G2 spot reports contributed considerably to rating omissions, the TOS must insure that sufficient format worksheets are on hand when it goes into the field.

The introduction of the automated TOS should considerably enhance the information available to the decision maker at all levels of command. While the implications for the TOS are all in the direction of increasing the proportion of message having certitude judgments, the question remains as to how to effect, concurrently, qualitative improvement in the evaluations. How can human performance, using the existing rating scales, be improved? Can better scales be developed to improve the quality of information provided the command decision maker?

Inclusion or exclusion of certitude measures is only one factor in a large and complicated operation. However, the consequence of failing to act on such small things was described by George Herbert in 1640 in Jacula Prudentum, when he began to describe the failure of a system in these words: "For want of a nail a shoe was lost . . ."

APPENDIX

THE INTELLIGENCE CYCLE AND THE NATO STANDARDIZED RATING SCALES FOR ASSIGNING RELIABILITY AND ACCURACY RATINGS TO SPOT REPORTS

THE INTELLIGENCE CYCLE AND THE NATO STANDARDIZED RATING SCALES FOR ASSIGNING RELIABILITY AND ACCURACY RATINGS TO SPOT REPORTS

Processing is the step in the intelligence cycle whereby information becomes intelligence. Processing consists of three basic operations.[1] The first of these operations is recording, which is reducing information to writing or some form of graphic representation and the grouping together of related items. The second operation is evaluation, which is a critical appraisal of information as a basis for its subsequent interpretation. The third operation is interpretation, which is the process of determining the significance of the information with respect to information and intelligence already known, and drawing conclusions as to the probable meaning of the evaluation information.

Figure A-1 shows how these processes enter into G2 operations at a division, corps, or field army headquarters.

While each of these operations is crucial, the present study was concerned only with the evaluation process, and, in fact, with a single aspect of evaluation--evaluation of the reliability of the information source and the accuracy of the information.

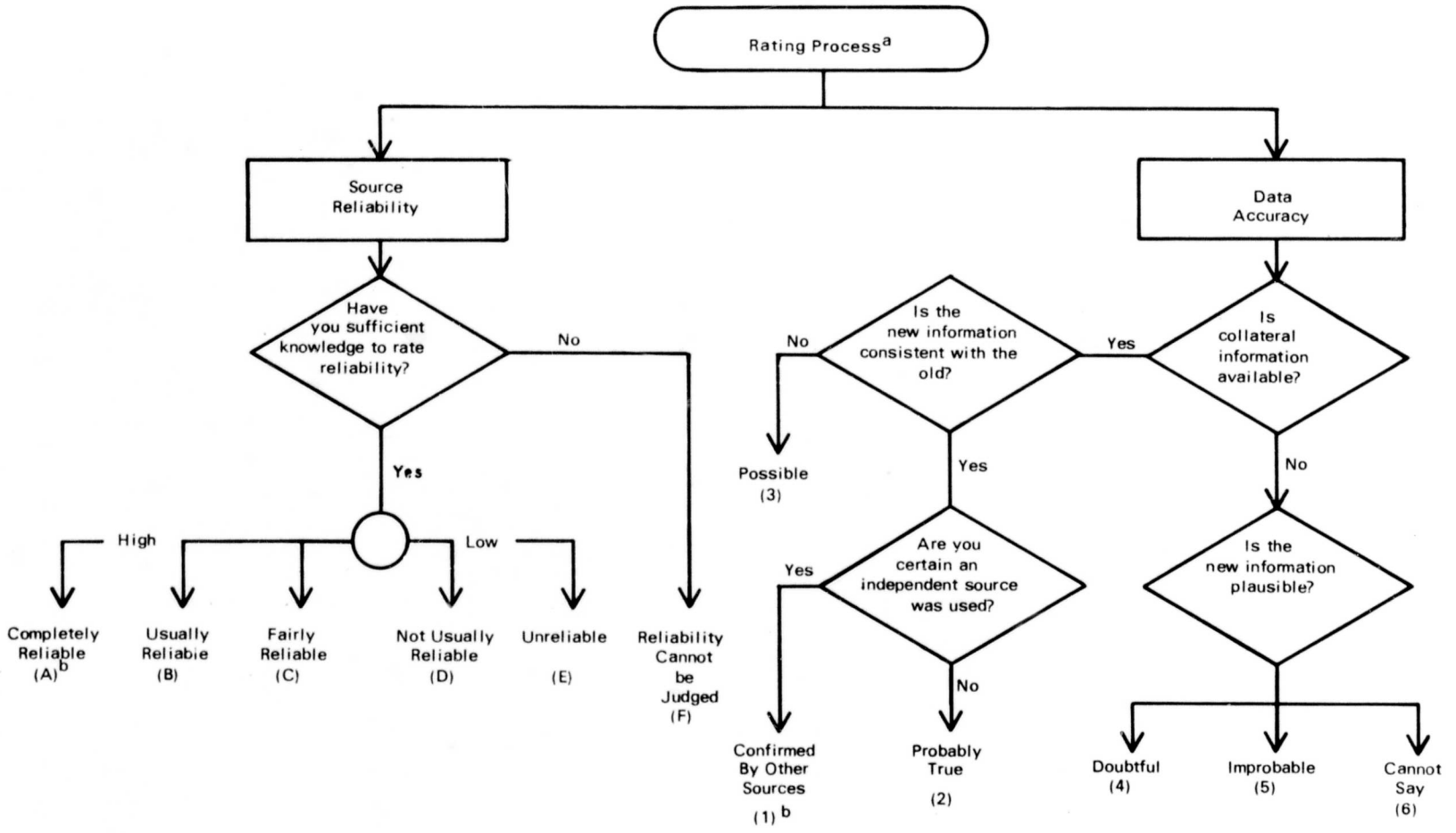
A system has been standardized for use by NATO standardized evaluation procedure for indicating the reliability of source and accuracy of information. This system utilized a letter to indicate reliability and a numeral for accuracy.

Evaluation ratings are made at the lowest headquarters possible; if information is incomplete, a partial evaluation rating may be given. Evaluation of the reliability of information source from which obtained is shown as follows:

A. Completely reliable
B. Usually reliable
C. Fairly reliable
D. Not usually reliable
E. Unreliable
F. Reliability cannot be judged

An A evaluation of source is assigned under only the most unusual circumstance, for example, when the source has long experience and extensive background with the type of information reported. The rating B indicates a source of known integrity. An F rating is assigned when there is no

[1] See Chapter 3 of Department of the Army Field Manual, FM 30-5, "Combat Intelligence", dated July 1963, for an extensive treatment of the concepts.



[a]Based on definitions and instructions in FM 30-5

[b]The letters in parentheses represent the value for the source reliability scale and the numbers in parentheses represent the numerical value for the data accuracy scale.

Figure 7. Decision table for determining source reliability and accuracy of information.

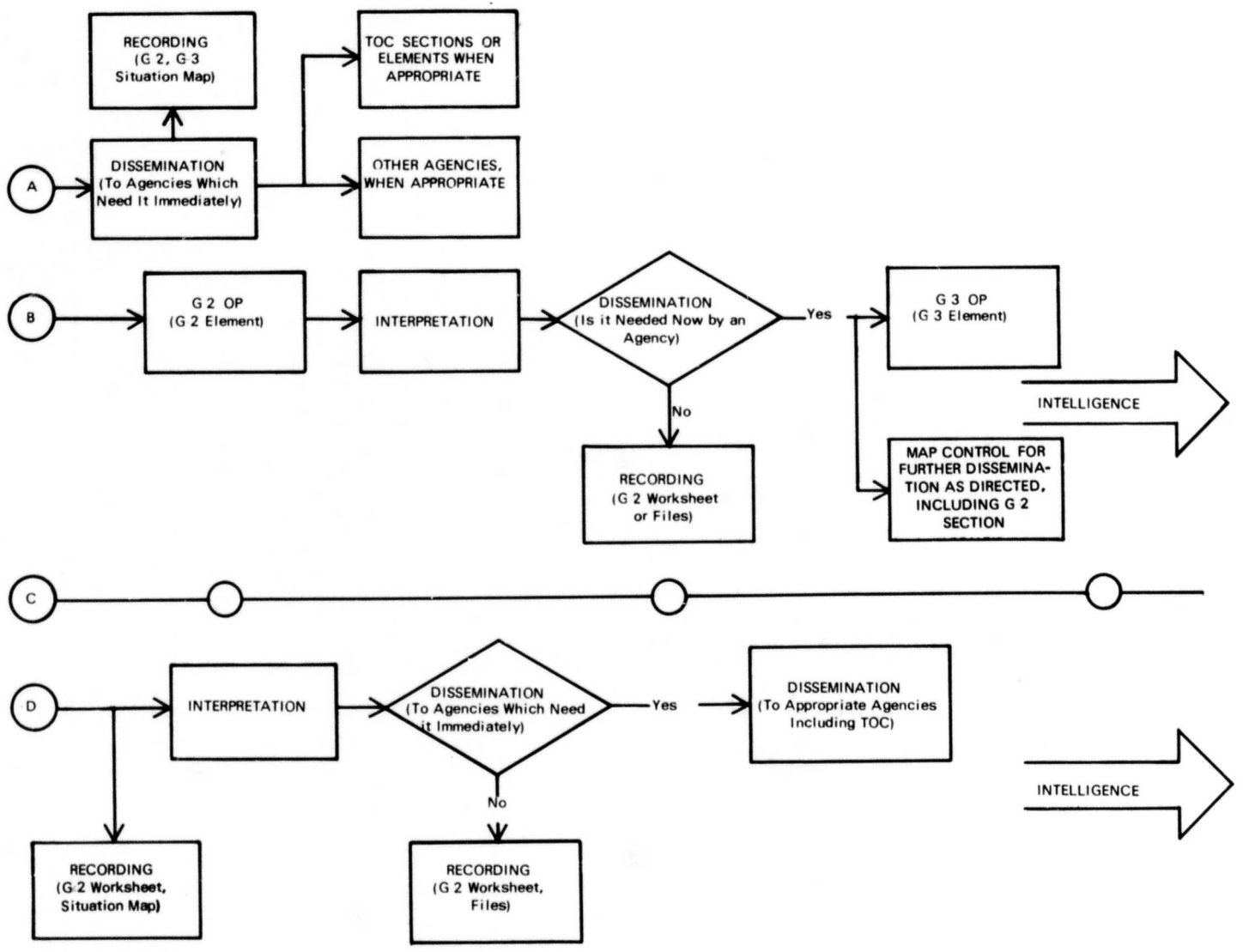


Figure A-1 (Continued)

adequate basis for estimating the reliability of the source. Agencies are ordinarily rated A. B, or C. However, when the source of an item and of the collecting-reporting agency are evaluated differently, only the lower degree of reliability is indicated.

Evaluation of the accuracy of an item of information is indicated as follows:

1. Confirmed by other sources
2. Probably true
3. Possibly true
4. Doubtfully true
5. Improbable
6. Truth cannot be judged

Determination of the values to be assigned is based on the considerations stated below.

Rating 1. Confirmed by other sources.

If it can be stated with certainty that the reported information originates from a source other than that for already existing information on the same subject, it will be classified as "confirmed by other sources" and will be rated "1".

Rating 2. Probably true.

If no proof in the above sense can be established, and if no reason exists to suspect that the reported information comes from the same source as the information already available on this subject, it will be classified as "probably true" and will be rated "2".

Rating 3. Possibly true.

If the information reveals that the reported facts--on which no further information is yet available--comply with behavior of the target as observed up to now, or if the known background of a person leads to the conclusion that he might have acted as reported, the information received will be classified as "possibly true" and will be rated "3".

Rating 4. Doubtfully true.

Reported but unconfirmed information the contents of which contradict the estimate of the development or the hitherto known behavior of the target will be classified as "doubtful" and will be rated "4" as long as this information cannot be disproved by available facts.

Rating 5. Improbable.

Reported information which is not confirmed by available data and which contradicts the experience hitherto assumed to be reliable with regard to the development of a case is classified as "improbable" and will be rated in category "5". The same classification is given to reported information that contradicts existing data on a subject originally rated "1" or "2".

Rating 6. Truth cannot be judged.

If the investigation of a report reveals that a basis for allocating ratings "1" to "5" is not given, the reported information will be classified as "truth cannot be judged" and will be rated "6". The statement that a report cannot be judged as to accuracy must always be preferred to an inaccurate use of the ratings "1" to "5". However, a rating "1" or "2" should always be tried; but if such a rating is not possible because of lack of other information on the same target, the rating "6" has to be given.

Although both letters and numerals are used to indicate the evaluation of an item of information, they are independent of each other. A completely reliable agency may report information obtained from a completely reliable source which, on the basis of other information, is judged to be improbable. In such a case, the evaluation of the information is A-5. A source known to be unreliable may provide information that is confirmed by other sources and is of undoubted accuracy. In such a case, a report is evaluated E-1. A report evaluated F-6 may be accurate and should not be arbitrarily discarded.

From the preceding description of the rating schema, it is evident that the decision process underlying the assignment of the alpha-numeric rating is complex. The complexity of this process may easily be overlooked when the output (an alpha-numeric value) alone is examined.

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)
U. S. Army Behavioral Science Research Laboratory, OCRD
HRB-Singer, Inc.

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP

3. REPORT TITLE

CERTITUDE JUDGMENTS IN AN OPERATIONAL ENVIRONMENT

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

5. AUTHOR(S) (First name, middle initial, last name)

James D. Baker (BESRL), James M. McKendry and Douglas J. Mace (HRB-Singer, Inc.)

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
November 1968	35	2

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO. DA R&D PJ No. 2J062106A723	Technical Research Note 200
c. TOS	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d. b-11	

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	AC of S for Force Development, ACSI, DA, Wash., D. C. Auto. Data Field Sys. Cmd., Ft. Belvoir

13. ABSTRACT

In September 1967, the Behavioral Science Research Laboratory (BESRL) established a Command Systems Field Branch in Europe to conduct human performance research in connection with the development of an experimental Tactical Operations System (TOS). The present publication describes one of the first studies conducted by the Field Branch. Major focus was placed upon the certitude judgment process within the TOS. In the intelligence cycle, G2 spot reports include subjective estimates of information expressed in a standard rating form. All messages (N = 2039) filed by two divisions of one corps during the 7-day field exercise were examined for presence of the required ratings of reliability of the source and accuracy of the information being sent. Assessment was made of the quantitative and qualitative improvement that could be expected under operating conditions within a computerized TOS. Observations during the field exercise revealed: 1) G2 spot reports constituted 70% of all messages; 2) Omission of the required source reliability and accuracy ratings occurred in 50% of the spot reports (attributed, in part, to unavailability of standard report forms); 3) Reliability and accuracy ratings were noticeably interdependent (distributions of the values indicated that senders tended to assign high scale values to ratings); 4) Analysis of the rating scales revealed the source reliability scale to be unidimensional and the accuracy scale to be multidimensional. Findings point to substantial increase in quantity of certitude data available for decision making at all levels of command by introduction of the automated TOS.

DD FORM 1 NOV 65 1473 REPLACES DD FORM 1473, 1 JAN 64, WHICH IS OBSOLETE FOR ARMY USE.

Unclassified
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
*Decision making						
Information assimilation						
Display modalities						
*Tactical Operations System (TOS)						
*Certitude judgments						
*Information display						
Command information processing systems						
*Intelligence information evaluation						
*Intelligence operations cycle						
Information storage and retrieval						
*Automated TOS						